Data-driven approach to identify field-scale biogeochemical transitions using geochemical and geophysical data and hidden Markov models: Development and application at a uranium-contaminated aquifer

Jinsong Chen,[1] Susan S. Hubbard,[1] and Kenneth H. Williams[1]

[1] Although mechanistic reaction networks have been developed to quantify the biogeochemical evolution of subsurface systems associated with bioremediation, it is difficult in practice to quantify the onset and distribution of these transitions at the field scale using commonly collected wellbore datasets. As an alternative approach to the mechanistic methods, we develop a data-driven, statistical model to identify biogeochemical transitions using various time-lapse aqueous geochemical data (e.g., Fe(II), sulfate, sulfide, acetate, and uranium concentrations) and induced polarization (IP) data. We assume that the biogeochemical transitions can be classified as several dominant states that correspond to redox transitions and test the method at a uranium-contaminated site. The relationships between the geophysical observations and geochemical time series vary depending upon the unknown underlying redox status, which is modeled as a hidden Markov random field. We estimate unknown parameters by maximizing the joint likelihood function using the maximization-expectation algorithm. The case study results show that when considered together aqueous geochemical data and IP imaginary conductivity provide a key diagnostic signature of biogeochemical stages. The developed method provides useful information for evaluating the effectiveness of bioremediation, such as the probability of being in specific redox stages following biostimulation where desirable pathways (e.g., uranium removal) are more highly favored. The use of geophysical data in the approach advances the possibility of using noninvasive methods to monitor critical biogeochemical system stages and transitions remotely and over field relevant scales (e.g., from square meters to several hectares).

1. Introduction

[2] One of main challenges with in situ remediation approaches is the difficulty in quantifying key states and transitions that are diagnostic of the efficacy of the treatment using conventional measurement and interpretation methods. Many efforts have been made to monitor and understand the changes of the systems at a range of scales and using various approaches. For example, *Nico et al.* [2009] and *Armstrong and Ajo-Franklin* [2011] developed high-resolution imaging techniques to visualize small-scale physical changes due to bioremediation. *Wan et al.* [2011]

developed column-scale experiments to identify biogeochemical dynamics. *Yabusaki et al.* [2007], *Fang et al.* [2009], and *Li et al.* [2010] developed reactive transport models to identify main reactive pathways or networks at the laboratory and field scales.

[3] Geophysical techniques, particularly induced-polarization (IP) methods, have been shown to be useful for monitoring changes in saturated, porous media caused by biostimulation. Several recent studies have documented the change in IP signatures associated with remediation treatments, many of which have been associated with biostimulation experiments conducted at the US Department of Energy (DOE) Rifle Integrated Field Research Challenge (IFRC) site near Rifle, Colorado (USA). For example, *Williams et al.* [2005], *Ntarlagiannis et al.* [2005], *Davis et al.* [2006], *Slater et al.* [2007], and *Wu et al.* [2011] used column-scale laboratory experiments to show how complex resistivity varies with specific processes induced through bioremediation treatments. *Williams et al.* [2009, 2011] and *Flores Orozco et al.* [2011] used surface IP data to track subsurface changes associated with field-scale remediation experiments.

---

[1]Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA.

Corresponding author: J. Chen, Earth Sciences Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA. (jchen@lbl.gov)

[4] Biostimulation studies conducted at the aforementioned Rifle field site have repeatedly shown that two predominant microbial metabolic pathways accompany acetate injection into the alluvial aquifer: iron-reduction followed by sulfate reduction [*Anderson et al.*, 2003; *Vrionis et al.*, 2005; *Yabusaki et al.*, 2007; *Fang et al.*, 2009; *Li et al.*, 2010; *Williams et al.*, 2011]. Injection of acetate acting as an organic carbon source and electron donor initially stimulates fast growing iron-reducing bacteria (e.g., members of the *Geobacteraceae*) commonly found in the Rifle sediments [*Williams et al.*, 2011]. Iron reduction involves the reductive dissolution of iron (hydr)oxide minerals to ferrous iron Fe(II); at the Rifle IFRC, this process is typically concurrent with the reductive immobilization of aqueous U(VI) to insoluble U(IV). With sustained acetate injection, respiration of sulfate by sulfate reducing bacteria leads to an accumulation of aqueous S(−II) and precipitation of sparingly soluble FeS given abundant Fe(II) produced during iron reduction.

[5] However, the state transitions from iron-reduction to sulfate-reduction under field conditions at the site are complicated because the spatial distribution and availability of ferric iron are unknown a priori and the spatial distribution of acetate concentrations is affected by physical and geochemical heterogeneity. *Williams et al.* [2011] and *Flores Orozco et al.* [2011] qualitatively divided the field-scale biogeochemical processes into four stages during bioremediation experiments according to borehole geochemical measurements and showed that these four stages were consistent with IP responses based on previous laboratory biogeophysical experiments. The first stage represents the period when iron reduction is the dominant form of metabolism and the second stage represents the period where both iron and sulfate reduction are concurrent forms of metabolism. The third stage represents the period when sulfate reduction is the predominant metabolic pathway in terms of sulfate consumption where concentrations decrease from preamendment levels of 10 m*M* to <1 m*M*. The fourth stage represents the period following cessation of acetate injection and the postinjection recovery of the system with significantly lower levels of stimulated microbial activity given exhaustion of acetate. Important for the uranium immobilization at the Rifle IFRC site, different reactive networks drive each of the stages, with each having characteristic biological (and abiotic) processes associated with it. The goal of this study is to develop an effective method to quantitatively identify those state transitions using time-lapse surface geophysical and borehole geochemical measurements.

[6] Hidden Markov or closely related Markov-switching models are effective approaches for identifying underlying states in space-time processes by integrating multisource and multiscale information [*Zucchini and MacDonald*, 2009]. They have been used for decades in earth and environmental sciences to estimate rainfall and runoff by combining large-scale atmospheric information with local-scale or regional-scale hydrological measurements. For example, *Zucchini and Guttorp* [1991] developed a hidden Markov model to estimate space-time rainfall by assuming the existence of unobservable climate states. *Hughes and Guttorp* [1994] developed a nonhomogeneous hidden Markov model to estimate multistation rainfall by using unobserved weather states to relate synoptic atmospheric patterns to regional hydrologic phenomena. *Lu and Berliner* [1999] developed a Markov-switching model to estimate daily runoff series by considering hydrological processes as three different stages (i.e., rising, falling, and normal phases), each of which can be modeled differently.

[7] In this study, we develop a hidden Markov model to quantitatively identify geochemical stages using borehole geochemical measurements and surface IP data. Both data sets were collected at multiple time points during an acetate-amendment experiment at the Rifle IFRC site. We consider the underlying stages as hidden states, with temporal transitions among those states driven by underlying reaction networks, the injection of amendments, and the local biogeochemical environment. We assume that each stage has a unique suite of geochemical characteristics defined by the probability distribution of the multivariate geochemical measurements. We stress that the intent of this study is not to mechanistically define petrophysical relationships between IP responses and specific geochemical transformations as has been developed by many previous studies using Rifle-based data or samples, but instead to develop a methodology that jointly uses time-lapse geochemical and geophysical data to identify integrated geochemical stages (such as predominance of a given redox condition) and their transitions over time. However, we build upon the petrophysical understanding developed through the previous studies.

[8] The remainder of this paper is organized as follows. Section 2 briefly describes the Rifle IFRC site and data used for the analysis, which provide the basis for developing the hidden Markov model. Section 3 describes the development of the hidden Markov model. The estimation results are given in section 4 and discussion and conclusions are provided in sections 5.

2. Rifle Biostimulation Experiments and Data Sets

2.1. Rifle IFRC Site and In Situ Bioremediation

[9] Several field-scale bioremediation experiments have been conducted at the uranium contaminated DOE Rifle IFRC site near Rifle, Colorado (USA) from 2002 to 2009. A detailed description of the site and experiments can be found in *Anderson et al.* [2003], *Vrionis et al.* [2005], and [*Williams et al.*, 2011]. The shallow subsurface consists of an unconfined, uranium-contaminated alluvial aquifer that includes sandy-gravely unconsolidated sediments with variable silt and clay content. Underlying the aquifer is a relatively impermeable aquitard (i.e., silt and mudstones of the Eocene Wasatch formation) located at spatially variable depths of 5.9–7.0 m below ground surface [*Williams et al.*, 2011; *Chen et al.*, 2012]. During the field experiments, acetate as an electron donor and organic carbon source was injected into the groundwater through a series of injection wells, along with the conservative tracer bromide.

[10] This study focuses on the biostimulation field experiments that were conducted between August 2007 and December 2009; the timeline of different amendment injections and details on the acquisition of geophysical measurements are given in *Flores Orozco et al.* [2011]. Specifically, we focus on the geochemical and geophysical
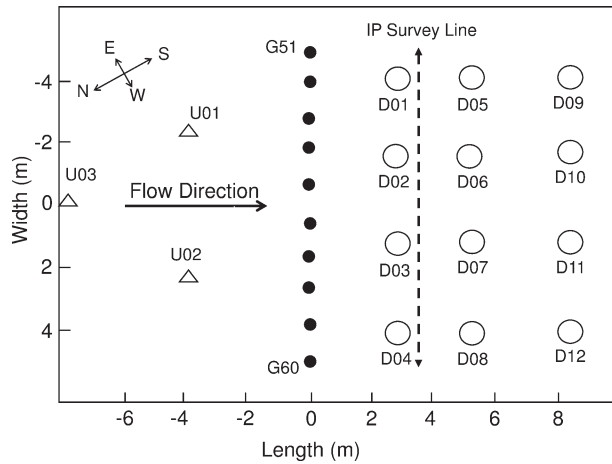
Figure 1. Schematic plan view of the 2007–2009 bioremediation experiments. The solid circles (G51–G60) are the 10 injection boreholes and the open circles (D01–D12) are the 12 down-gradient monitoring wells. The open triangles (U01–U03) are the three up-gradient monitoring wells, and the dashed line is the survey profile of surface induced polarization used for this study.

Figure 2. Acetate concentrations as a function of days after the initial experiment. Day 0 of the biostimulation experiment is 8 August 2007, and the black, red, green, and blue curves show the data collected from boreholes D1, D2, D3, and D4, respectively.

data collected from 19 July 2008 to 8 December 2009. Figure 1 illustrates the well field used to conduct the bioremediation experiments, where the 10 solid circles (i.e., G51–G60) are acetate injection boreholes and the 12 open circles (i.e., D01–D12) are down-gradient monitoring wells. Acetate was injected into the unconfined aquifer over the saturated interval of 3.5–6.0 m below ground surface. The three open triangles (i.e., U01–U03) in Figure 1 are up-gradient monitoring wells. Time-lapse surface induced polarization data were collected along the dashed line, which is located at 2.7 m down gradient from the injection wells. Geochemical sampling and geophysical data collection (both described below in detail) occurred before, during, and after the period of acetate injection.

2.2. Borehole Aqueous Geochemical Measurements

[11] Under the Rifle field conditions, multiple reactions (e.g., microbial and geochemical) and multiple processes (i.e., sorption, desorption, dissolution, etc.) may occur during acetate amendment [*Li et al*, 2010]. It is difficult to isolate the reactions and processes and to measure the microbial and geochemical components separately. Instead, it is common to infer the reactions based on data available from groundwater samples.

[12] For this study, we collected groundwater samples from the depth of 5 m below ground surface from each of the 12 down-gradient monitoring wells as a function of time after injection of acetate, with the temporal sampling intervals of two or three days. We performed geochemical analysis and measured multiple components of aqueous geochemical concentrations, including Fe(II), sulfate, sulfide, acetate, uranium, chloride, and bromide concentrations. The fluid samples are representative of the groundwater conditions at depths approximately 0.15 m above and below the discrete sampling locations.

[13] Figure 2 shows the acetate concentrations at boreholes D1 (black), D2 (red), D3 (green), and D4 (blue) as a
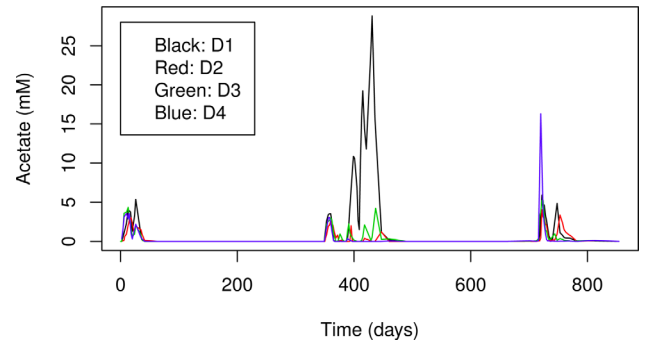
function of elapsed days after the initial experiment starting on 8 August 2007. This figure shows that the acetate concentrations in borehole D1 in the second experiment period around Day 400 (i.e., from 31 August 2008 to 17 November 2008) are significantly higher than those in other boreholes and during other periods. Since our focus in the current study is on the identification of redox states under acetate-based biostimulation, we only work on the data collected in borehole D1 from Days 389 to 483 because both geochemical and geophysical data are available on the period. Figure 3 shows logarithmic acetate, bromide, sulfate, sulfide, Fe(II), and uranium concentrations as a function of elapsed time during the period of interest.

2.3. Surface Induced Polarization Data

[14] We collected surface induced polarization data along the profile near boreholes D1, D2, D3, and D4, referred to as Array A and shown as the dashed line in Figure 1. The array consisted of 30 electrodes with 1 m spacing, leading to a profile length of 29 m. The impedance measurements were carried out using a dipole-dipole configuration with each dipole skipping three electrodes, resulting in a dipole length of 4 m for the current and potential dipoles [*Flores Orozco et al.*, 2011]. The recorded impedance data were first inverted for complex resistivity using the least squares based algorithms developed by *Kemna* [2000] and *Binley and Kemna* [2005]. The inverted results were then converted to complex conductivity and subsequently used for the current study. The measurement protocol was used in a previous study, where it was demonstrated to produce good resolution data for depths to 7 m [*Williams et al.*, 2009]. In this study, we use five different frequencies (i.e., 0.25, 0.5, 1, 2, and 4 Hz) to collect IP data from Day 361 to Day 853 after the beginning of acetate injection. More methodological details regarding the data acquisition are presented in *Flores Orozco et al.* [2011].

[15] Figure 4 shows the real and imaginary conductivity data collected from borehole D1 for frequencies 0.25 Hz (black), 0.5 Hz (red), 1 Hz (green), 2 Hz (blue), and 4 Hz (cyan) as a function of elapsed days. Since the IP data from different frequencies have similar trends, we use only those obtained at frequency 0.5 Hz for this study. We expect that the choice of which frequency to include in the estimation
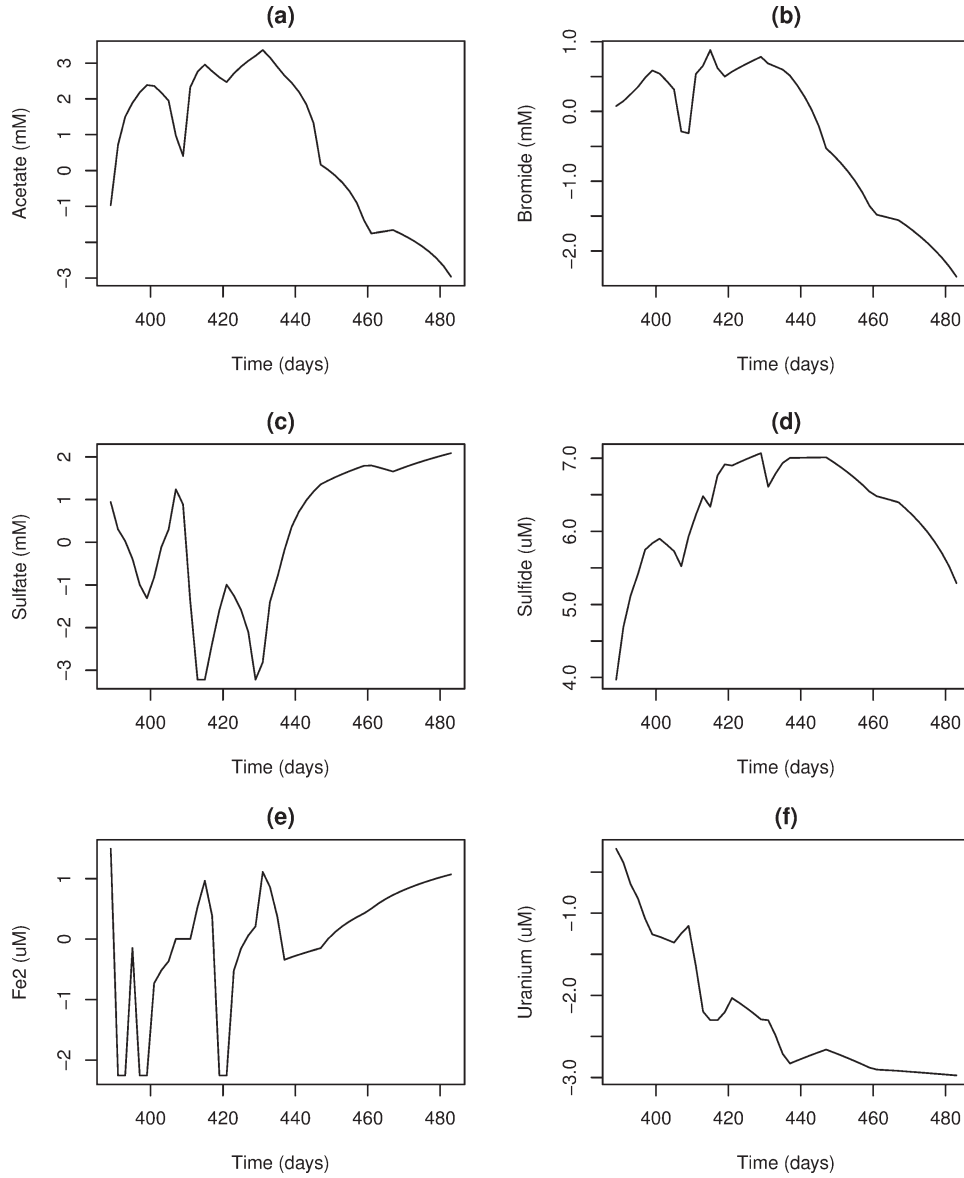
Figure 3. Logarithmic geochemical data collected from borehole D1: (a) acetate (mM), (b) bromide (mM), (c) sulfate (mM), (d) sulfide (μM), (e) Fe(II) (μM), and (f) uranium (μM) as a function of the elapsed time.

will have a limited effect on the geochemical stage estimation, because as will be subsequently discussed, we use regression-based models that mainly rely on the trends of the IP responses (i.e., increasing or decreasing), which are similar for all frequencies.
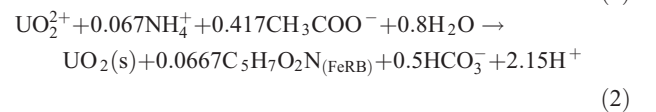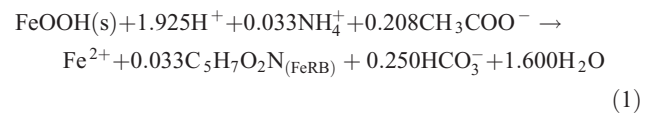
3. Hidden Markov Model

[16] We describe development of the hidden Markov model for estimating the transitions of the underlying geochemical stages during bioremediation based on the main reaction networks identified from the Rifle IFRC site. We begin by introducing the main reaction networks and then describe the data-driven model. The aqueous geochemical concentrations used for the development include acetate, bromide, sulfate, sulfide, Fe(II), and uranium concentrations as shown in Figure 3; the geophysical data include

the real and imaginary components of IP data at frequency 0.5 Hz as a function of elapsed time.

3.1. Microbe-Mediated Reaction Networks and Data-Driven Hidden Markov Model

[17] Although many reactions occur at the Rifle site, the system responses to acetate-based biostimulation can be represented using the subset of reactions shown in Figure 5 and provided by *Li et al.* [2010]:

$$FeOOH(s)+1.925H^{+}+0.033NH_4^{+}+0.208CH_3COO^{-} \rightarrow$$
$$Fe^{2+}+0.033C_5H_7O_2N_{(FeRB)} + 0.250HCO_3^{-}+1.600H_2O$$
(1)

$$UO_2^{2+}+0.067NH_4^{+}+0.417CH_3COO^{-}+0.8H_2O \rightarrow$$
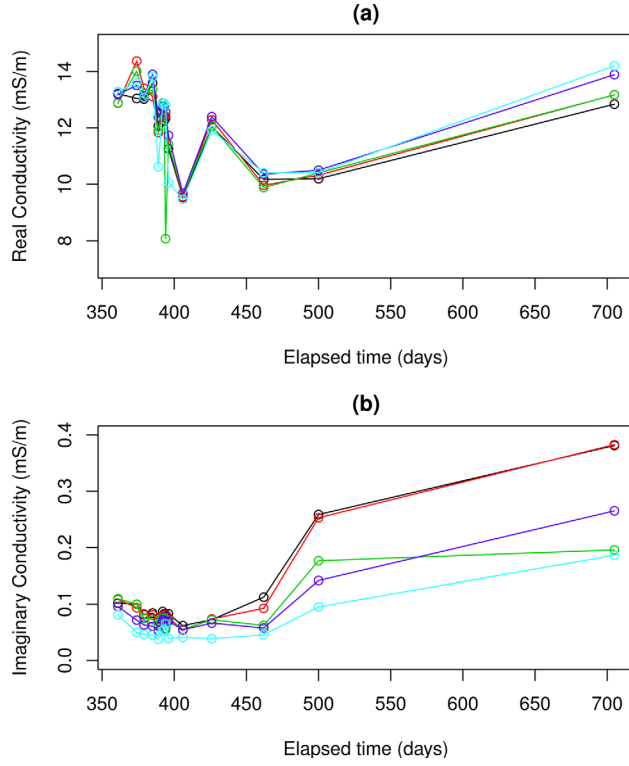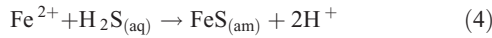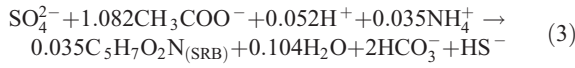$$UO_2(s)+0.0667C_5H_7O_2N_{(FeRB)}+0.5HCO_3^{-}+2.15H^{+}$$
(2)

Figure 4. Induced polarization data at borehole D1: (a) real conductivity (mS/m) and (b) imaginary conductivity (mS/m) at frequencies 0.25 Hz (black), 0.5 Hz (red), 1 Hz (green), 2 Hz (blue), and 4 Hz (cyan) as a function of elapsed days.

$$SO_4^{2-}+1.082CH_3COO^-+0.052H^++0.035NH_4^+ \rightarrow$$
$$0.035C_5H_7O_2N_{(SRB)}+0.104H_2O+2HCO_3^-+HS^- \quad (3)$$

$$Fe^{2+}+H_2S_{(aq)} \rightarrow FeS_{(am)} + 2H^+ \quad (4)$$

[18] Formulas 1 and 2 are the main reactions for iron reduction. Accompanying the activity of iron reducing bacteria (i.e., FeRB), insoluble Fe3+ is reduced to aqueous Fe2+ and soluble U6+ is reduced to solid phase U4+. The third formula is the main reaction of sulfate reduction accompanying the activity of sulfate reducing bacteria (i.e., SRB). The last formula is the precipitation of new metal mineral induced by metabolic end-products of both FeRB
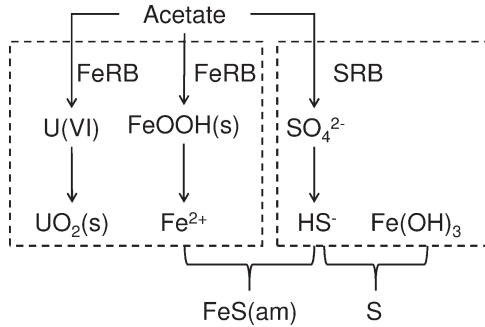
and SRB; it is believed to be the main cause of geophysical responses.

[19] At the Rifle IFRC site, several studies [*Yabusaki et al.*, 2007; *Fang et al.*, 2009; *Li et al.*, 2010] have developed numerical reactive transport models, with the use of above main reaction paths and some other geochemical and geophysical processes, to simulate the field-scale bioremediation processes. Given the spatial distribution of hydrogeological and geochemical parameters (e.g., permeability and bioavailable Fe(III)) and under suitable initial and boundary conditions, those models reproduce the borehole aqueous groundwater geochemical responses well. The mechanistic approach is critical for understanding the fundamentals of field-scale bioremediation processes. However, those methods are computationally expensive and subject to large degree of uncertainty because both model calibration and associated parameter estimation are complex inverse problems.

[20] In this study, we take a complementary but different approach to the mechanistic modeling methods. We adopt a data-driven, statistical approach, which can incorporate geophysical data and other types of available information. We develop a statistical model to identify the unknown geochemical state transitions defined by the main reaction networks using measured aqueous geochemical and geophysical data. The underlying principle for such development lies in the fact that each stage of the processes has unique geochemical characteristics as shown in Formulas 1–4. Rather than utilizing a mechanistic representation of the reaction networks (see Formulas 1–4), we develop a new approach based on hidden Markov models for quantifying critical biogeochemical transitions within the complex subsurface system.

[21] Figure 6 illustrates the general structure of the hidden Markov model developed for this study, where the arrows show dependent relationships. In the figure, S_1, S_2, S_3, ..., and S_T represent state variables taking integers between 1 and m, where m is the total number of possible states and T is the total number of time steps for estimation. For example, value 1 corresponds to iron reduction state and value 2 corresponds to sulfate reduction state. All those state variables are unknown and hidden to us. Vectors X_1, X_2, X_3, ..., and X_T represent aqueous geochemical measurements (e.g., Fe(II), sulfate, sulfide, etc.) and geophysical data. It is reasonable to assume that the observed data depend in some ways on their corresponding unknown states. In the estimation procedure, we can develop linkages between X_1 and X_2, X_2 and X_3, ..., and so on if we

Figure 5. Main reaction networks associated with the Rifle uranium biostimulation experiments.
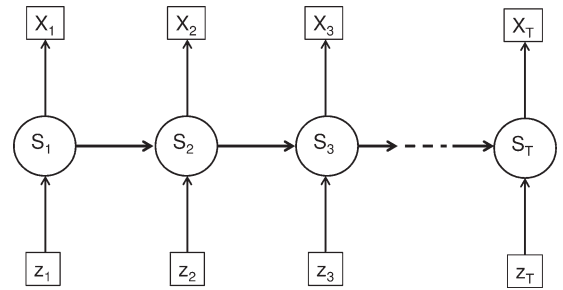

Figure 6. General structure of the developed hidden Markov model.

want to account for the temporal dependence of geochemical and geophysical parameters. The temporal changes of underlying states form a Markov chain, which is inhomogeneous because the state transitions may depend on other types of information.

3.2. State-Dependent Geophysical and Geochemical Characteristics

[22] Many models can be used to describe the dependence of geophysical and geochemical data on the underlying states and they are typically multivariate. In this study, we build state-dependent regression models based on geochemical Formulas 1–4 and the availability of geophysical and geochemical measurements. For geochemical data, we consider acetate and bromide as exploratory variables since acetate is an injected electron donor for iron and sulfate reduction (see Formulas 1 and 3) and bromide is a conservative tracer that is coinjected during the field experiments. We also consider sulfate as an exploratory variable since it is reduced to sulfide during sulfate reduction (see Formula 3). We consider Fe2+, uranium, and sulfide as dependent variables because Fe2+ and uranium are altered by iron reduction and sulfide is the product of sulfate reduction. For IP data, since solid phase FeS is the primary cause of IP responses [*Pelton et al.*, 1978] and by Formula 4, FeS is formed from the chemical reaction of Fe2+ and sulfide, we consider IP data as dependent variables of Fe2+ and sulfide. Consequently, we have the following regression equations:

$$Fe2 \sim \alpha_{0i} + \alpha_{1i}\text{Acetate} + \alpha_{2i}\text{Sulfate} + \alpha_{3i}\text{Sulfate} \times \text{Acetate}$$
$$+ \alpha_{4i}\text{Bromide} + \varepsilon_{1i}, \tag{5}$$

$$U \sim \beta_{0i} + \beta_{1i}\text{Acetate} + \beta_{2i}\text{Bromide} + \varepsilon_{2i}, \tag{6}$$

$$Sulfide \sim \theta_{0i} + \theta_{1i}\text{Acetate} + \theta_{2i}\text{Sulfate} + \theta_{3i}\text{Sulfate} \times \text{Acetate}$$
$$+ \theta_{4i}\text{Bromide} + \varepsilon_{3i}, \tag{7}$$

$$IPreal \sim u_{0i} + u_{1i}\text{Fe2} + u_{2i}\text{Sulfide} + u_{3i}\text{Fe2} \times \text{Sulfide} + \varepsilon_{4i}. \tag{8}$$

$$IPimag \sim v_{0i} + v_{1i}\text{Fe2} + v_{2i}\text{Sulfide} + v_{3i}\text{Fe2} \times \text{Sulfide} + \varepsilon_{5i}. \tag{9}$$

[23] Symbols α_{ki}, β_{ki}, θ_{ki}, u_{ki}, and v_{ki} are unknown coefficients of regression equations and ε_{ki} represents normally distributed random errors with unknown standard deviation, where $k = 0$, 1, 2, 3 or 4 and i represents a state. Terms "Fe2," "Acetate," "Sulfate," "Bromide," "U," and "Sulfide" represent the logarithmic concentrations of aqueous Fe2+, acetate, sulfate, bromide, U6+, and sulfide, respectively. The terms "IPreal" and "IPimag" represent the real and imaginary conductivity data collected at frequency 0.5 Hz.

[24] Ideally, the change in acetate concentrations caused by the reactions should be used in the estimation procedure rather than the total acetate concentration values invoked in the above regression equations. However, it is easier to measure acetate concentrations at wellbores than to measure their changes. To compensate for the effects of transport on the acetate concentration, we include bromide concentrations as a variable in equations (5)–(7). The necessity of using the bromide term will be determined by model selection procedures, as will be described below. Although the microbe-mediated reaction shown in Formula 1 does not include sulfate, we include two terms related to sulfate in equation (5) as Fe2+ may react with sulfide during iron reduction (see Formula 4). Again, the necessity of those terms in the estimation procedure will be determined by model selection. To take account for possible interactions between acetate and sulfate, we included their product as a regression term. In equations (5)–(9), we do not include Fe3+ and U4+ because both are not commonly assayed through groundwater-based measurements given their insoluble state.

[25] All the coefficients and standard deviations of residuals in the regression equations are unknown, and they depend on the status of redox-based transformations. We estimate those coefficients using parameter estimation and model selection procedures described below.

3.3. State Evolution and Transition Probability Models

[26] We use a Markov model to describe the underlying redox-based state transitions associated with acetate-based bioremediation at the Rifle site. Since the total number of states often is more than two, we use a multinomial distribution to describe them. Let γ_{ij} represent the transition probability from the ith state to the jth state. The transition probability may depend on other types of categorical information, such as the injection status of acetate z_t at time t, and it is a conditional probability function given below:

$$\gamma_{ij}(z_t) = P(S_{t+1} = j | S_t = i, \ z_t). \tag{10}$$

[27] We may divide acetate concentrations into two categories (i.e., low or high status) or three categories (i.e., low, medium, or high status). For general, let p be the total number of the categories. We let variable z_t be a vector of size p with its component being 1 if a particular status (i.e., low-acetate, medium-acetate, or high-acetate concentrations) presents and 0 otherwise. Consequently, we can use a multinomial logistic model [*Venables and Ripley*, 1999] to obtain the transition probability as follows

$$\gamma_{ij}(z_t) = \frac{\exp\left(z_t^T w_j\right)}{1 + \sum_{k \neq j*} \exp\left(z_t^T w_k\right)}, \tag{11}$$

where w_i and w_k are the coefficient vectors of size p that correspond to the current state $S_t = i$ and $S_t = k$, respectively. The letter T is the transpose of a vector and j^* is the arbitrary baseline redox state. For the baseline category, the transition probability is given by:

$$\gamma_{ij*}(z_t) = 1 - \sum_{k \neq j*} \gamma_{ik}(z_t). \tag{12}$$

3.4. Likelihood Function and Parameter Estimation

[28] We follow the methods and notations given by *Zucchini and MacDonald* [2009] to develop the likelihood

function for parameter estimation. Let $p_i(x) = \Pr(\mathbf{X}_t = x | S_t = i)$ be the conditional probability of $\mathbf{X}_t$ when the Markov chain is in the state i at time t. The likelihood function is a joint conditional probability distribution of all the data given all the unknown parameters, which can be expressed as:

$$L_T = \Pr(x_1, x_2, \cdots, x_T | \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{u}, \mathbf{v}) = \boldsymbol{\delta} \mathbf{P}(x_1) \prod_{t=2}^{T} \boldsymbol{\Gamma}(z_t) \mathbf{P}(x_t) \mathbf{U},$$

(13)

where $\mathbf{U} = (1, 1, \cdots, 1)^T$, $\boldsymbol{\delta} = (\delta_1, \delta_2, \cdots, \delta_m)^T$, the initial probability of each unknown state, and $\mathbf{P}(x_t)$ is a diagonal matrix whose elements are $p_1(x_t)$, $p_2(x_t)$, ..., $p_m(x_t)$. The symbol $\boldsymbol{\Gamma}(z_t)$ is the transition matrix at time step t, which is given below:

$$\boldsymbol{\Gamma}(z_t) = \begin{pmatrix} \gamma_{11}(z_t) & \gamma_{12}(z_t) & \cdots & \gamma_{1m}(z_t) \\ \gamma_{21}(z_t) & \gamma_{22}(z_t) & \cdots & \gamma_{2m}(z_t) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m1}(z_t) & \gamma_{m2}(z_t) & \cdots & \gamma_{mm}(z_t) \end{pmatrix}.$$

(14)

[29] We use the Expectation-Maximization algorithm [*Dempster et al.*, 1977] to estimate unknown parameters by maximizing the likelihood function given in equation (13). The algorithm is an iterative method that finds maximum likelihood estimates of parameters when some of the data are missing by following two main steps (i.e., the Expectation-step and Maximization-step). In the Expectation-step, we compute the conditional expectation of the missing data given the observations and given the current estimates of parameters. In the Maximization-step, we maximize, with respect to parameters, the complete data log likelihood with the functions of the missing data replaced in it by their conditional expectation. The detailed algorithms for the hidden Markov model are given in *Zucchini and MacDonald* [2009]. For implementation, we use an R package, called "depmixS4," developed by *Visser and Speekenbrink* [2010] for hidden Markov models. *Visser* [2011] provides a good tutorial on several key issues in hidden Markov modeling.

3.5. Model Selection

[30] Since the underlying bioremediation processes are very complex, we explore a range of models in terms of state-dependent probability distributions to determine the best model for estimation. They include (1) different number of underlying states, (2) different numbers of regression equations, (3) various combinations of covariates, (4) various ways to use geophysical data, and (5) prior distribution of parameters. We use model selection techniques to select

the model that best represents the processes using minimal explanatory variables. In this study, we use the Bayesian information criterion (BIC) model selection technique developed by *Schwarz* [1978]:

$$\text{BIC} = -2 \log L + p \log T,$$

(15)

where $\log L$ is the log likelihood of the fitted model, p is the total number of unknown parameters in the model, and T is the total number of observations. The BIC often favors the models with fewer parameters than does the popular Akaike information criterion (AIC) [*Zucchini and MacDonald*, 2009].

4. Estimation Results

4.1. Parameter Estimation and Model Selection

[31] Based on what we consider to be the "full" model representations given by equations (5)–(9), we explore a variety of submodels based on assumptions about which data are available for the estimation algorithm. To be concise, we only describe the results of following five cases: (1) use of geochemical data only, (2) use of geochemical and real conductivity data, (3) use of geochemical and imaginary conductivity data, (4) use of geochemical and real and imaginary conductivity data (i.e., "full model"), and (5) use of geochemical data (excluding bromide) and imaginary conductivity data. For each combination of the model setup, we use the expectation-maximization algorithm to estimate parameters and calculate log likelihood and their corresponding BIC values.

[32] A summary of the model selection results is given in Table 1. The total number of unknown parameters showing in Column 2 is calculated based on three states. The log likelihoods shown in the last column generally decrease with the decreasing of the number of unknown parameters because a model with more parameters typically fits data better for the same model setting. However, the Bayesian information criterion (BIC) does not always increase with the decreasing total number of unknown parameters since it is punished by the total number of unknowns.

[33] The smallest BIC should give us the best model for the given data under the model family that we specify. Under this criteria, the best model is the one that uses geochemical data plus imaginary conductivity because it gives the BIC value of 35.61, which is significantly smaller than the BIC value of using geochemical data only (BIC = 94.63), using both geochemical and real conductivity data (BIC = 113.89), or using geochemical data plus both real and imaginary conductivity data (BIC = 110.31). This means that the field imaginary conductivity data

Table 1. A Summary of Model Selection Results Using the Bayesian Information Criterion

Cases	Total Number of Unknowns	Bayesian Information Criterion (BIC)	Log Likelihood
Case 1: Geochemical data only (equations (5)–(7))	56	94.63	61.08
Case 2: Geochemical and IP real conductivity data (equations (5)–(8))	71	113.89	80.48
Case 3: Geochemical and IP imaginary conductivity data (equations (5)–(7), (9))	71	35.61	119.62
Case 4: Geochemical and IP real and imaginary conductivity data (equations (5)–(9))	86	110.31	111.31
Case 5: Geochemical (excluding bromide) and IP imaginary conductivity data (equations (5)–(7), (9), excluding bromide terms)	62	102.96	68.53

provide significant information about the underlying redox states. To explore the necessity of including bromide as an exploratory variable, we exclude all the bromide terms from the best model. This leads to the BIC value of 102.96, which is significantly larger than that of the best model.

4.2. Interpretation of the Identified Underlying State Transformations

[34] The approach allows us to estimate the coefficients of equations (5)–(9), thereby developing the relationships between geochemical and geophysical parameters for critical redox-based states. Table 2 summarizes the results of the estimation procedure. We interpret the results based on the relative magnitudes of coefficients and not the signs of the coefficients because each component is related to multiple reaction pathways and potentially multiple physical processes (e.g., sorption and desorption).

[35] From Table 2, we can see that at each of the identified states, all the reactions given in Formulas 1–4 have occurred. This means no pure iron reduction or sulfate reduction phase exists under the field conditions, unlike in the laboratory column experiments. However, at different states, there are different dominant reactions, which allow us to distinguish between critical states. For example, from Table 2, we can see the regression equation (i.e., equation (9)) of imaginary conductivity versus Fe2 and sulfide for state 3 is very different from that in states 1 and 2. For state 3, the IP responses are strongly related to Fe2, sulfide, and their product. This is because Fe2 and sulfide together form solid phase FeS, which in turn affects the IP response. For states 1 and 2, the IP responses are weakly related to Fe2, sulfide, and their product. We also examine the regression equation of sulfide versus acetate and sulfate because this is one of the main features of sulfate reduction. As shown in Table 2, the sulfide in state 1 is more closely related to sulfate and acetate compared to state 2. On the other hand, uranium in state 2 is more closely related to acetate concentration than that in state 1. Considering the various reaction pathways, we call states 1, 2, and 3 sulfate reduction "dominated," iron reduction "dominated," and recovery states, respectively; we refer them to as "SRB," "IRB," and "REC" for convenience.

[36] Figure 7 compares the identified three states with corresponding geochemical and geophysical data. This figure shows that when both acetate and sulfide concentrations are high, Fe2 concentrations are very low. This is possibly because the abundance of sulfide produced by sulfate reduction converts Fe2 to FeS. In the recovery phase, acetate decreases and both sulfate and Fe2 rebound.

[37] We can extract transition probabilities among the three identified states from the estimated best model, which can provide information about field-scale bioremediation processes. The following is the transition probability matrix obtained from the best model.

$$\begin{pmatrix} 0.500 & 0.500 & 0.000 \\ 0.187 & 0.750 & 0.063 \\ 0.000 & 0.000 & 1.000 \end{pmatrix} \qquad (16)$$

[38] In the above matrix, the first, second, and third rows and columns correspond to the states of sulfate reduction (SRB), iron reduction (IRB), and the recovery state (REC), respectively. From the matrix, we can see that there are strong transitions between the states of sulfate-dominated and iron-dominated reduction.

4.3. State-Dependent Relationships Among Fe(II), Uranium, and Acetate

[39] We examine correlations among Fe(II), uranium (U6+), and acetate concentrations as they are closely related to iron reduction as shown in Figure 5 and Formulas 1 and 2. Figure 8 shows the crossplots of Fe(II), uranium, and acetate concentrations as a function of underlying states; their correlation coefficients are given in Table 3. As given in the last row of Table 3, the overall correlations among those geochemical concentrations are low (i.e., less than or around 0.5). For example, the correlation coefficients of Fe(II) and uranium with acetate are −0.4766 and 0.4112, and the correlation coefficient between Fe(II) and uranium is −0.5299. However, with the identification of underlying states, we can get much better correlations among those parameters.

[40] For the iron reduction dominated state (see the crosses in Figure 8 and the third row in Table 3), we can

Table 2. Interpretation of the Estimated States Based on the Relative Magnitudes of Coefficients in the Regression Equations (Excluding Constant Terms)

| States	Fitted Models
State-1 (sulfate reduction dominated state, referred to as "SRB")	Fe2 ∼ −0.591 × Acetate + 1.233 × Sulfate + 1.512 × Acetate × Sulfate − 1.377 × Bromide (sd = 0.036) U ∼ −1.287 × Acetate + 0.036 × Bromide (sd = 0.416) Sulfide ∼ 1.671 × Acetate − 1.384 × Sulfate − 3.660 × Acetate × Sulfate + 0.297 × Bromide (sd = 0.302) IPimag ∼ 0.626 × Fe2 − 0.565 × Sulfide + 0.460 × Fe2 × Sulfide (sd = 0.260)
State-2 (iron reduction dominated state, referred to as "IRB")	Fe2 ∼ 2.211 × Acetate + 7.893 × Sulfate + 4.731 × Acetate × Sulfate − 5.401 × Bromide (sd = 0.312) U ∼ −3.810 × Acetate + 2.662 × Bromide (sd = 0.318) Sulfide ∼ 3.779 × Acetate − 0.049 × Sulfate − 2.157 × Acetate × Sulfate − 0.225 × Bromide (sd = 0.407) IPimag ∼ −0.100 × Fe2 + 0.500 × Sulfide + 0.281 × Fe2 × Sulfide (sd = 0.093)
State-3 (recovery state, referred to as "REC")	Fe2 ∼ 0.303 × Acetate − 0.934 × Sulfate − 1.487 × Acetate × Sulfate + 0.336 × Bromide (sd = 0.045) U ∼ −0.318 × Acetate + 0.447 × Bromide (sd = 0.067) Sulfide ∼ −2.309 × Acetate − 0.764 × Sulfate + 1.900 × Acetate × Sulfate + 1.212 × Bromide (sd = 0.069) IPimag ∼ 4.099 × Fe2 + 2.331 × Sulfide − 2.352 × Fe2 × Sulfide (sd = 0.074)
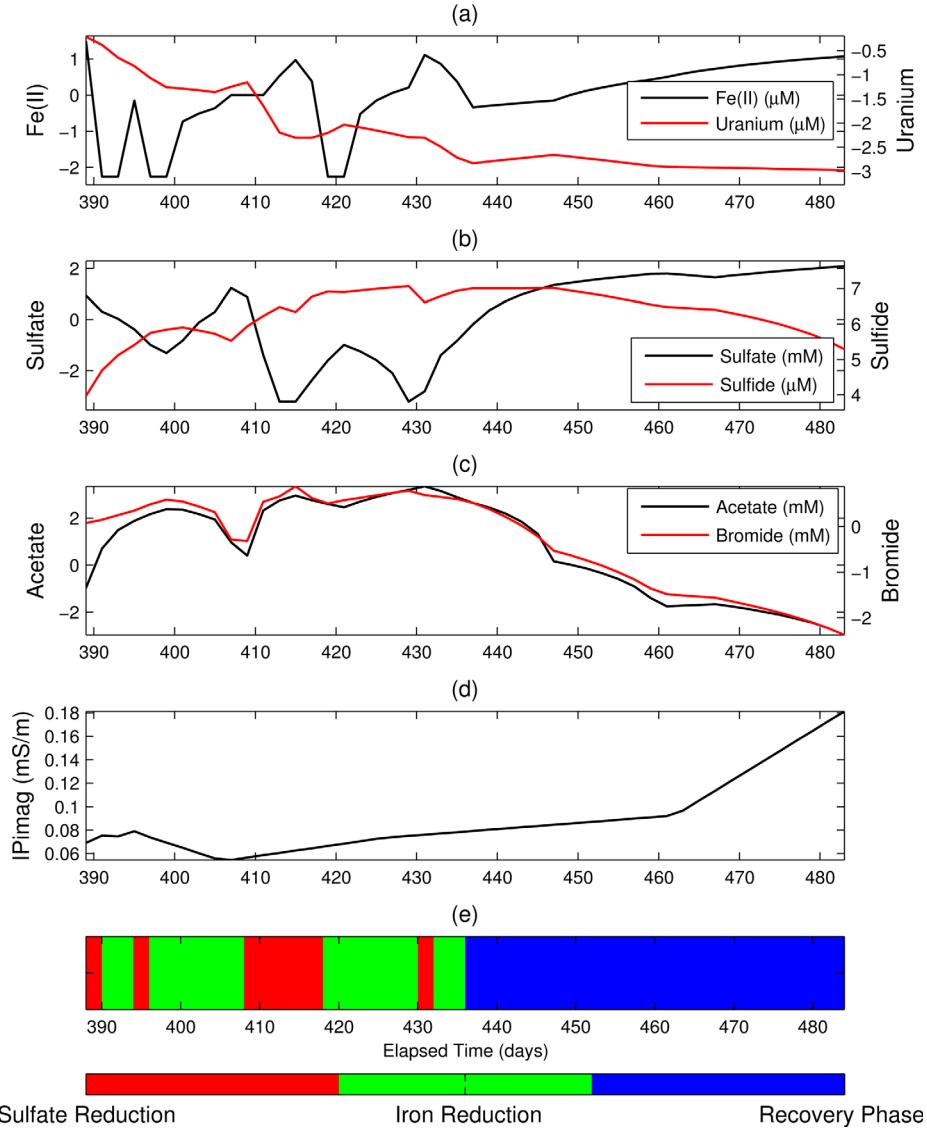
Figure 7. Comparison of estimated states and their corresponding geochemical and geophysical data.

see that (1) Fe(II) concentrations generally increase with increasing acetate concentrations (Corr = 0.4568), (2) uranium concentrations decrease with increasing acetate concentrations (Corr = −0.8649), and (3) Fe(II) and uranium concentrations are negatively correlated (Corr = −0.5958). Those results are consistent with reaction Formulas 1–2 because Fe(II) is the product of the reduction of bioavailable Fe3+, and aqueous U6+ is transformed to insoluble U4+ accompanying the iron reduction.

[41] For the sulfate reduction dominated state (see the circles in Figure 8 and the second row in Table 3), we can see that uranium varies over a large range (see Figure 8c). The high negative correlation (Corr = −0.9015) between uranium and acetate concentrations (see Figure 8b) means that the decreasing of acetate corresponds to the increasing of uranium. For the recovery state (see the triangles and the fourth row in Table 3), Fe(II) concentrations are observed to increase with decreasing of acetate concentrations (see the triangles in Figure 8a, and Corr = −0.9580), opposite to the iron reduction dominated state (see the crosses in

Figure 8a). This is because at this state no iron reduction occurs and the consuming of acetate will produce more sulfides that transform Fe(II) to FeS.

4.4. State-Dependent Relationships Among Sulfate, Sulfide, and Acetate

[42] Next, we examine correlations among sulfate, sulfide, and acetate concentrations because they are closely related to sulfate reduction as shown in Figure 5 and Formula 3. Figure 9 shows the crossplots of sulfate, sulfide, and acetate concentrations as a function of underlying states; their detailed correlation coefficients are given in Table 4. From the last row of Table 4, we can see that sulfate and acetate concentrations have a very strong negative correlation (Corr = −0.8687), but the overall correlations of acetate and sulfate with sulfide are low (i.e., 0.3426 and −0.2241, respectively).

[43] For the sulfate reduction dominated state (see the circles in Figure 9 and the second row in Table 4), sulfate is negatively correlated to acetate (Corr = −0.9134) and
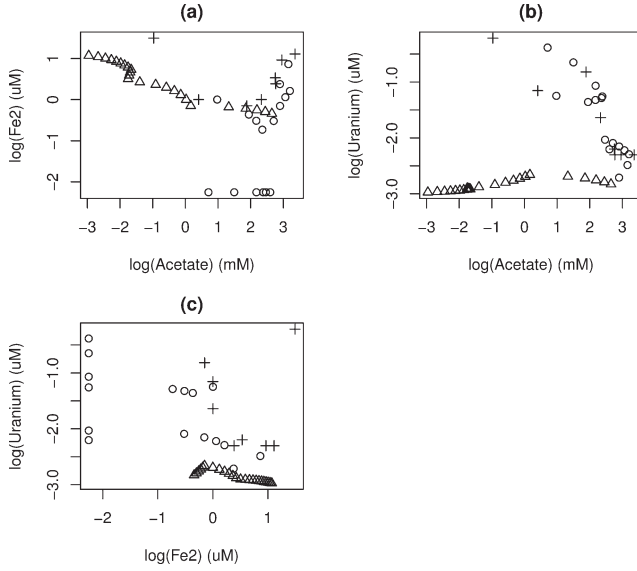
Figure 8. Crossplots of Fe(II), uranium, and acetate concentrations as a function of estimated states, where the crosses, circles, and triangles represent the iron reduction, sulfate reduction, and recovery states, respectively.

Figure 9. Crossplots of acetate, sulfate, and sulfide concentrations as a function of estimated states, where the crosses, circles, and triangles represent the iron reduction, sulfate reduction, and recovery states, respectively.

sulfide is positively correlated to acetate (Corr = 0.8845) as sulfate reduction consumes both sulfate and acetate to produce sulfide (Formula 3). We also found that sulfide has a strong negative correlation with sulfate (Corr = −0.7745), as expected stoichiometrically.

[44] We see similar patterns for the iron reduction dominated state (see the crosses in Figure 9 and the third row in Table 4) and for the recovery state (see the triangles and the fourth row in Table 4). This is because under the field conditions, we cannot clearly separate each state, and sulfate reduction (albeit at low levels) may occur in both iron reduction dominated [*Druhan et al.*, 2012] and recovery states as shown in Table 2.

4.5. State-Dependent Relationships Between Imaginary Conductivity and Fe(II) and Sulfide

[45] We examine relationships among imaginary conductivity and Fe(II) and sulfide concentrations because Fe(II) and sulfide together can form FeS (see Formula 4) that in turn affects the IP response. Figure 10 shows the crossplots of imaginary conductivity, Fe(II), and sulfide concentrations as a function of underlying states; their correlation coefficients are given in Table 5. The overall correlations among
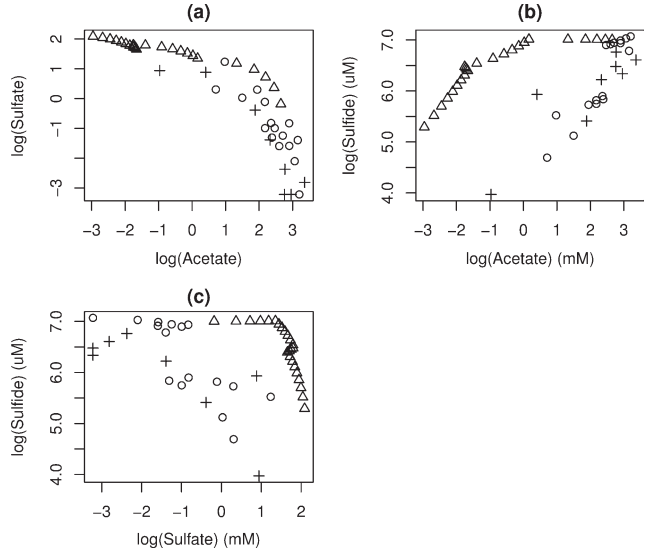
imaginary conductivity, Fe(II), and sulfide are quite low (Corr (IPimag, Fe(II)) = 0.4798, Corr (IPimag, sulfide) = −0.1951, and Corr (Fe(II), sulfide) = 0.0214).

[46] For the recovery state, Figure 10a shows the IP responses increase with increasing of Fe(II) (Corr = 0.8742). This is likely because the Fe(II) recovery happens to have the same trend as the iron precipitation (FeS). Figure 10b shows the IP responses increase with decreasing sulfide concentrations. This suggests that the reaction given in Formula 4 is primarily controlled by the availability of sulfide given the abundance of Fe(II) (see Figure 7a) and FeS precipitation increases with the consuming of sulfide. From Figure 10c, we can see that to have large IP responses, we need relatively high concentrations of both Fe(II) and sulfide. From this figure, we can see for some iron reduction dominated states (see the crosses on the top right corner), both Fe(II) and sulfide concentrations are high. Those possibly lead to the precipitation of FeS (see Figure 7d around Day 400) in a short period, but the persistent IP responses need accumulation of FeS over a certain time period. For the both iron and sulfate reduction dominated states (crosses and circles), as shown in Figure 10, the IP responses are very low.

Table 3. State-Dependent Pairwise Correlations of Fe(II), Uranium, and Acetate Concentrations

	Corr (Fe(II), Acetate)	Corr (Uranium, Acetate)	Corr (Fe(II), Uranium)
Sulfate reduction dominated state	−0.1624	−0.9015	−0.0079
Iron reduction dominated state	0.4568	−0.8649	−0.5958
Recovery state	−0.9580	0.7785	−0.8936
All states	−0.4766	0.4112	−0.5299

Table 4. State-Dependent Pairwise Correlations of Sulfide, Sulfate, and Acetate Concentrations

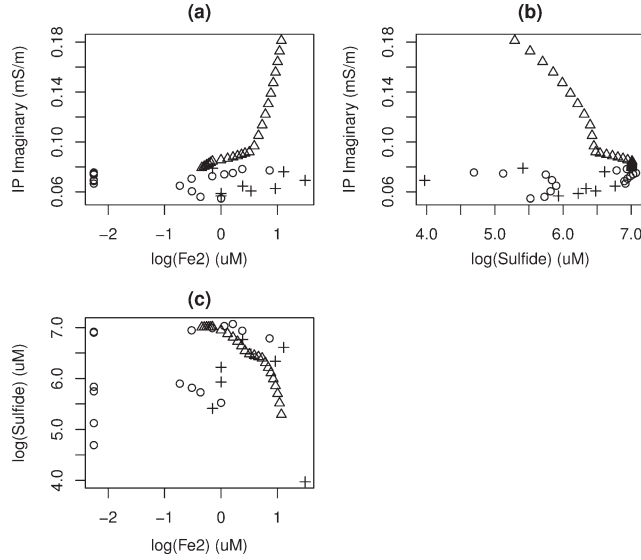|	Corr (Sulfate, Acetate)	Corr (Sulfide, Acetate)	Corr (Sulfate, Sulfide)
Sulfate reduction dominated state	−0.9134	0.8845	−0.7745
Iron reduction dominated state	−0.8567	0.9027	−0.7617
Recovery state	−0.9338	0.8525	−0.7134
All states	−0.8687	0.3426	−0.2241

Figure 10. Crossplots of Fe(II) and sulfide concentrations and IP imaginary conductivity data as a function of estimated states, where the crosses, circles, and triangles represent the iron reduction, sulfate reduction, and recovery states, respectively.

5. Discussion and Conclusions

[47] We adopted a data-driven approach in this study to classify the field-scale biogeochemical transitions by using the time-lapse borehole aqueous geochemical measurements and surface IP data under several assumptions. First, we made a key assumption that the aqueous geochemical concentrations are mediated by microbial reactions and not from the flow and reactive transport. In reality, the measured concentrations are influenced by both transport and local reactions. Since all the geochemical components used in the state-dependent equations are local and contemporaneous and we include bromide concentrations as covariates in the regression equations, we think this assumption is reasonable. Second, the estimated results are subject to uncertainty because of the model setup and the method for finding solutions. For example, we consider only four relationships in the state-dependent equations, and we may need to add more equations to account for more processes in the bioremediation processes. In the current model, we ignore the temporal correlation of geochemical parameters themselves and assume that all the temporal dependence comes from the hidden states. Direct incorporation of the temporal correlation may change the detailed results to

Table 5. State-Dependent Pairwise Correlations of IP Imaginary Conductivity and Fe(II), and Sulfide Concentrations

	Corr (IPimag, Fe(II))	Corr (IPimag, Sulfide)	Corr (Fe(II), Sulfide)
Sulfate reduction dominated state	0.2237	−0.2620	−0.3251
Iron reduction dominated state	−0.0197	0.2899	0.4614
Recovery state	0.8742	−0.9720	−0.9299
All states	0.4798	−0.1951	0.0214

some degrees, but the main conclusions are expected to be same.

[48] Mechanistic approaches (i.e., both reactive transport and geophysical models) can be very helpful for improving the results of the current studies. For instance, reactive transport modeling can take account of a wide range of chemical reactions besides the main pathways in Formulas 1–4 and the spatial heterogeneity of physical, geochemical, and microbial parameters [*Li et al.*, 2010]. Such simulation will pose constraints on multiple geochemical parameters that we use for estimating underlying redox states. Similarly, mechanistic petrophysical models, such as electrochemical models for disseminated sulfide ores [*Wong*, 1979] and mechanistic models for shaly sands [*Revil*, 2012], could be incorporated to link IP responses directly to electrolytes and their activities. Development and incorporation of mechanistic petrophysical models, which may reduce uncertainty or ambiguity between IP responses and geochemical parameters, warrants further investigation.

[49] Our results from the Rifle field data sets are generally consistent with those found from previous laboratory column experiments based on samples collected at the same site [*Williams et al.*, 2009], but some differences exist. Under complex field conditions, the transition between iron and sulfate reduction states is not as clear cut as it is in the laboratory experiments. At the field scale, iron reduction and sulfate reduction appears to occur concurrently, which is in agreement with *Williams et al.* [2011]; their relative dominance depends on the local availability of acetate, sulfate, bioavailable Fe3+, and so on. Consequently, we refer to the identified states to as (1) the sulfate reduction dominated state, (2) the iron reduction dominated state, and (3) the recovery state.

[50] At borehole D1, we identify that the sulfate reduction dominated state occurs first, followed by the iron reduction dominated state for the selected time period from Day 390 to Day 480. This is different from laboratory column experiments using fresh Rifle sediments (i.e., previously unstimulated by acetate) [*Williams et al.*, 2005], where iron reduction starts first followed by sulfate reduction. But it is in general agreement with the field observations of *Druhan et al.* [2012] who documented the early onset of sulfate reduction in the same experimental plot presented here, where previously stimulated sediments (in 2007) are predisposed toward sulfate reduction and rapid onset following secondary stimulation in the same plot. Another possible reason for the sequence is that iron reduction occurs before Day 390 as we can see from Figure 2 during the earliest period of acetate injection (noting that acetate injection and arrival of acetate at D1 proceeded day 390).

[51] We examine cross correlations among various geochemical data and their association with IP responses. Although overall pairwise correlations among different geochemical data are low (i.e., less than or around 0.5, excluding sulfate versus acetate), the new approach yields greatly improved state-dependent correlations (greater than 0.6) for those pairs that are mechanistically connected. Particularly, the pairwise correlations of Fe(II) and sulfide with IP imaginary conductivity for the recovery state have been significantly improved (from less than 0.5 to more than 0.87).

[52] The identified state-dependent relations provide strong field-scale evidence to support the IP mechanisms associated with bioremediation found by many previous laboratory column experiments. For example, *Williams et al.* [2005] and *Ntarlagiannis et al.* [2005] using Rifle sediments showed that the IP anomalies are caused by microbe-induced metal sulfide precipitation (FeS and ZnS) because they change the electrical charge transport through the sediments by encouraging a flow of dissolved ions to and from the metal-electrolyte interface, which causes an excess or deficit of inactive ions to accumulate there [*Wong*, 1979]. *Slater et al.* [2007] and *Personna et al.* [2008] further showed through lab experiments that the IP responses associated with bioremediation of Rifle sediments likely result from the formation of FeS biofilms on the mineral surface of pores in the sand matrix, not the biominerals encrusted bacterial cells themselves, and that the IP signatures were fully reversible under anaerobic and aerobic conditions due to the ephemeral nature of the bioprecipitates. The strong correlations of imaginary conductivity with Fe(II) (Corr = 0.87) and with sulfide (Corr = −0.97) in the recovery stage, and the need to maintain relatively high concentrations for both Fe(II) and sulfide (see Figure 10c) suggest that the above-mentioned IP mechanisms found from laboratory experiments likely also contribute to the observed field scale IP responses.

[53] In addition, column-scale laboratory studies also demonstrated the microbially mediated IP responses could be reduced because of other concurrent processes, such as the formation of biofilm polymers [*Williams et al.*, 2005], the decrease of total surface areas due to coagulation of individual biominerals [*Slater et al.*, 2007; *Chen et al.*, 2009], and the precipitation of calcite [*Wu et al.*, 2009]. Our results (see Figure 10c) show that the IP response is small at the sulfate reduction stage, where both Fe(II) and sulfide have high concentrations. As suggested by *Flores Orozco et al.* [2011], among many possible reasons, the concurrent formation of calcite during acetate injection [*Li et al.*, 2010] may depress the IP response during that stage. Compared to the column experiments that were performed to investigate IP mechanisms at the Rifle site, the goal of this study is to identify the states and transitions that most influence the IP response. It is not our intent to identify the associated mechanisms at each stage that lead to the effective IP response, and indeed the strength of our method is that it provides information about the system transitions without requiring such knowledge. However, interpretations based on our data-driven approach are well aligned with previous laboratory study findings.

[54] The developed method provides a wide range of quantitative information for understanding in situ reaction pathways and thus the complex processes associated with bioremediation. We have obtained dynamic relationships among various geochemical concentrations as a function of the underlying redox states, geochemical and geophysical features as a function of the underlying state, and the sequences of state transitions under the field condition. Those results may be used to constrain field-scale reactive transport modeling, for example, by providing lower and upper bounds or best estimated values for some parameters, and determining which pathways dominate under local environments.

[55] Although we estimated the state transitions as a function of time at only one borehole location in this study, the method can be extended to incorporate 2-D IP profile information and multiple wellbore data sets. For the Rifle site, such extension would require consideration of the sequences of the transition states along each of the four boreholes and use of a hierarchical Bayesian model similar to the one by *Chen et al.* [2012] to estimate the spatiotemporal distribution of underlying states by conditioning to data at boreholes and geophysical data along the profile. Such an extension is currently in progress and is expected to lead to valuable information to constrain or compare with reactive transport simulations, and to generally understand remediation-induced processes over field relevant space and time scales.

[56] Testing of our developed methodology at the Rifle site suggests that individual geochemical or geophysical measurements alone could not provide sufficient information on the field-scale bioremediation processes. This means individual geochemical or geophysical data are not solely sufficient as diagnostic indicators of field-scale bioremediation-based transitions at the Rifle IFRC site. However, the identified redox states, which is obtainable using geochemical and geophysical time-lapse data sets, take account of multiple reaction pathways and thus serve as an integrated signature that is diagnostic of field-scale systems-level transitions and feedbacks. The developed data-driven methodology offers a unique avenue for incorporating diverse data sets to improve our predictive understanding of complex, dynamic subsurface systems.

References

Akaike, H. (1973), Information theory and an extension of the maximum likelihood principle, in *Second International Symposium on Information Theory*, edited by B. N. Petrov and F. Csaki, pp. 267–281, Akademiai Kiado, Budapest.

Anderson, R. T., et al. (2003), Stimulating the in-situ activity of *Geobacter* species to remove uranium from the groundwater of a uranium-contaminated aquifer, *Appl. Environ. Microbiol.*, *69*(10), 5884–5891.

Armstrong, R., and J. Ajo-Franklin (2011), Investigating biomineralization using synchrotron based X-ray computed microtomography, *Geophys. Res. Lett.*, *38*, L08406, doi:10.1029/2011GL046916.

Binley, A., and A. Kemna (2005), DC resistivity and induced polarization methods, in *Hydrogeophysics*, edited by Y. Rubin and S. Hubbard, pp. 129–156, Springer, Dordrecht, Netherlands.

Chen, J., S. Hubbard, K. Williams, S. Pride, L. Li, C. Steefel, and L. Slater (2009), A state-space Bayesian framework for estimating biogeochemical transformations using time-lapse geophysical data, *Water Resour. Res.*, *45*, W08420, doi:10.1029/2008WR007698.

Chen, J., S. S. Hubbard, K. H. Williams, A. Flores-Orozco, and A. Kemna (2012), Estimating the spatiotemporal distribution of geochemical parameters associated with biostimulation using spectral induced polarization data and hierarchical Bayesian models, *Water Resour. Res.*, *48*, W05555, doi:10.1029/2011WR010992.

Davis, C. A., E. Atekwana, E. Atekwana, L. D. Slater, S. Rossbach, and M. R. Mormile (2006), Microbial growth and biofilm formulation in geologic media is detected with complex conductivity measurements, *Geophys. Res. Lett.*, *33*, L18403, doi:10.1029/2006GL027312.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977), Maximum likelihood from incomplete data via the EM algorithm (with discussion), *J. R. Stat. Soc., Ser. B*, *39*, 1–38.

Druhan, J. L., C. I. Steefel, S. Molins, K. H. Williams, M. E. Conrad, and D. J. DePaolo (2012), Timing the onset of sulfate reduction over multiple subsurface acetate amendments by measurement and modeling of sulfur isotope fractionation, *Environ. Sci. Technol.*, *46*(16), 8895–8902.

Fang, Y., S. B. Yabusaki, S. J. Morrison, J. P. Amonette, and P. E. Long (2009), Multicomponent reactive transport modeling of uranium bioremediation field experiments, *Geochim. Cosmochim. Acta*, *73*(20), 6029–6051.

Flores Orozco, A., K. H. Williams, P. E. Long, S. S. Hubbard, and A. Kemna (2011), Using complex resistivity imaging to infer biogeochemical processes associated with bioremediation of an uranium contaminated aquifer, *J. Geophys. Res.*, *116*, G03001, doi:10.1029/2010JG001591.

Hughes, J. P., and P. Guttorp (1994), A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena, *Water Resour. Res.*, *30*, 1535–1546.

Kemna, A (2000), Tomographic inversion of complex resistivity: Theory and application, PhD Thesis, Ruhr-University Bochum, Bochum, Germany.

Li, L., C. I. Steefel, M. B. Kowalsky, A. Englert, and S. Hubbard (2010), Effects of physical and geochemical heterogeneities on mineral transformation and biomass accumulation during biostimulation experiments at Rifle, Colorado, *J. Contam. Hydrol.*, *112*, 45–63.

Lu, Z., and L. M. Berliner (1998), Markov switching time series models with application to daily runoff series, *Water Resour. Res.*, *35*(2), 523–534.

Nico, P. S., B. D. Stewart, and S. Fendorf (2009), Incorporation of oxidized uranium into Fe (Hydr)oxides during Fe(II) catalyzed remineralization, *Environ. Sci. Technol.*, *43*, 7391–7396.

Ntarlagiannis, D., K. H. Williams, L. Slater, and S. S. Hubbard (2005), Low-frequency electrical response to microbial induced sulfide precipitation, *J. Geophys. Res.*, *110*, G02009, doi:10.1029/2005JG000024.

Pelton, W. H., S. H. Ward, P. G. Hallof, W. R. Sill, and P. H. Nelson (1978), Mineral discrimination and removal of inductive coupling with multi-frequency IP, *Geophysics*, *43*(3), 588–609.

Personna, Y. R., D. Ntarlagiannis, L. Slater, N. Yee, M. O'Brien, and S. S. Hubbard (2008), Spectral induced polarization and electrodic potential monitoring of microbially mediated iron sulfide transformations, *J. Geophys. Res.*, *113*, G02020, doi:10.1029/2007JG000614.

Revil, A. (2012), Spectral induced polarization of shaly sands: Influence of electrical double layer, *Water Resour. Res.*, *48*, W02517, doi:10.1029/2011WR011260.

Schwarz, G. E. (1978), Estimating the dimension of a model, *Ann. Stat.*, *6*(2), 461–464.

Slater, L., D. Ntarlagiannis, Y. R. Personna, and S. S. Hubbard (2007), Pore-scale spectral induced polarization signatures associated with FeS biomineral transformations, *Geophys. Res. Lett.*, *34*, L21404, doi:10.1029/2007GL031840.

Venables, W. N., and B. D. Ripley (1999), *Modern Applied Statistics With S-Plus*, 3rd ed., Springer, New York.

Visser, I. (2011), Seven things to remember about hidden Markov models: A tutorial on Markovian models for time series, *J. Math. Psychol.*, *55*, 403–415.

Visser, I., and M. Speekenbrink (2010), depmixS4: An R package for hidden Markov models, *J. Stat. Software*, *36*(7), 1–21.

Vrionis, H. A., R. T. Anderson, I. Ortiz-Bernad, K. R. O'Neill, C. T. Resch, A. D. Peacock, R. Dayvault, D. C. White, P. E. Long, and D. R. Lovley (2005), Microbiological and geochemical heterogeneity in an in-situ uranium bioremediation field site, *Appl. Environ. Microbiol.*, *71*(10), 6308–6318.

Wan, J., W. Dong, and T. K. Tokunaga (2011), Method to attenuate U(VI) mobility in acidic waste plumes using humic acids, *Environ. Sci. Technol.*, *45*, 2331–2337.

Williams, K. H., D. Ntarlagiannis, L. Slater, A. Dohnalkova, S. S. Hubbard, and J. Banfield (2005), Geophysical imaging of stimulated microbial bioremediation, *Environ. Sci. Technol.*, *39*(19), 7592–7600.

Williams, K. H., A. Kemna, M. Wilkins, J. Druhan, E. Arntzen, L. N'Guessan, P. E. Long, S. S. Hubbard, and J. F. Banfield (2009), Geophysical monitoring of coupled microbial and geochemical processes during stimulated subsurface bioremediation, *Environ. Sci. Technol.*, *43*(17), 6717–6723.

Williams, K. H., et al. (2011), Acetate availability and its influence on sustainable bioremediation of uranium-contaminated groundwater, *Geomicrobiol. J.*, *28*, 519–539.

Wong, J. (1979), An electrochemical model of the induced-polarization phenomenon in disseminated sulfide ores, *Geophysics*, *44*(7), 1245–1265.

Wu, Y., R. Versteeg, L. Slater, and D. LaBrecque (2009), Calcite precipitation dominates the electrical signatures of zero valent iron columns under simulated field conditions, *J. Contam. Hydrol.*, *106*, 131–143.

Wu, Y., J. B. Ajo-Franklin, N. Spycher, S. S. Hubbard, G. Zhang, K. H. Williams, J. Taylor, Y. Fujita, and R. Smith (2011), Geophysical monitoring and reactive transport modeling of ureolytically-driven calcium carbonate precipitation, *Geochem. Trans.*, *12*, doi:10.1186/1467-4866-12-7.

Yabusaki, S. B., et al. (2007), Uranium removal from groundwater via in situ biostimulation: Field-scale modeling of transport and biological processes, *J. Contam. Hydrol.*, *93*, 216–235.

Zucchini, W., and P. Guttorp (1991), A hidden Markov model for space-time precipitation, *Water Resour. Res.*, *27*, 1917–1923.

Zucchini, W., and I. L. MacDonald (2009), *Hidden Markov Models for Time Series: An Introduction using R*, CRC Press, New York.